

Rethinking the typology of relative clauses

BERNARD COMRIE
University of Southern California

1. Introduction: The relative pronoun strategy¹

In the majority of the major literary languages of Europe, the main strategy for forming relative clauses is what has come to be called the relative pronoun strategy, illustrated for English in (1):

ENGLISH

(1) *the boy [whom I saw -]*

The main characteristics of this strategy are as follows: The position relativized is indicated inside the relative clause by means of a pronominal element which is preposed to the front of the relative clause, and which is case-marked to indicate its syntactic/semantic role within the relative clause. Thus, in (1), the relative pronoun *whom* is preposed-normally, direct objects follow their verbs in English-and is in the accusative case to indicate its syntactic role as direct object. Obviously, there are refinements that need to be made for a complete account, for instance in that the relative pronoun does not literally have to be the first word in its clause, but may take along other constituents such as prepositions when it is preposed, as in (2):

(2) *the boy [for whom I bought the book -]*

And as (2) also illustrates, "case marking" should be understood in a broad sense, to include the marking of syntactic/semantic roles by adpositions (prepositions or postpositions) as well as by morphological cases.

Examples (3)-(10) illustrate the relative pronoun strategy in a number of other European languages.

RUSSIAN

- (3) *mal' ik, kotorogo ja videl*
 boy which.ACC I saw
 'the boy [whom I saw -]'

GERMAN

- (4) *der Junge, den ich sah*
 the boy that.ACC I saw)
 'the boy [whom I saw -]'

RUSSIAN

- (5) *mal' ik sobaku kotorogo ja videl*
 boy dog.ACC who.GEN I saw)
 'the boy [whose dog I saw -]'

FRENCH

- (6) *le garçon dont j'ai vu le chien*
 the boy whose I have seen the dog)

HUNGARIAN

- (7) 'the boy [whose dog I saw -]'
a fiú, akit láttam
 the boy who.ACC I.saw

FINNISH

- (8) 'the boy [whom I saw -]'
poika, jonka näin
 boy REL.ACC. I.saw
 'the boy [whom I saw -]'

GEORGIAN

- (9) *c erili, romelic man c uxel dac era*
 letter.ABS REL.ABS he.ERG last.night wrote.3SG)
 'the letter [which he wrote - last night]'

ESPERANTO

- (10) *la...lingvo-j, kin-j-n oni trov-as en ĉinujo*
 the language.PL, which-PL-ACC one find-PRS in China)

'the languages [which one finds - in China]'

It should be noted that the relevant group of languages does really seem to be the areally defined group "European languages". While most of the languages included are Indo-European, some are not, such as the Uralic languages Finnish and Hungarian and the Kartvelian (South Caucasian) language Georgian. Also included is the perhaps most quintessential of all European languages, namely the artificial language Esperanto. As we will see below, Indo-European languages spoken outside Europe often have other strategies for relative clause formation. The more easterly Uralic languages also use different strategies, and the use of the relative pronoun strategy in Finnish and Hungarian is a clear reflection of areal contact with Germanic, Baltic, and Slavic languages. The areal typological relevance of the Georgian relative clause construction is noted by Shimomiya (1978).

Moreover, relative clauses formed using the relative pronoun strategy are quite exceptional outside Europe, except as a recent result of the influence of European languages, as we shall see below. The relative pronoun strategy thus seems to be a remarkable areal typological feature of European languages, especially the standard written languages. In the body of this article, I want to examine some ways in which this remarkable feature of European languages has led linguists to adopt a number of assumptions in their investigation of relative clauses in other languages, assumptions that have sometimes proved helpful, but which in other cases have arguably led linguists astray.

But before turning to this point, one final clarification should be added. The mere presence of a pronoun that is restricted to relative clauses, and is thus in some intuitive sense a relative pronoun, is not sufficient to define an instance of the relative pronoun strategy. As example (11) from Modern Standard Arabic shows, such a relative pronoun can be case-marked not to indicate its role in the relative clause, but rather to agree in case with the head noun in the matrix clause; in (11) the relative pronoun is nominative, like the head noun, whereas the position relativized in the relative clause is direct object (which would require the accusative case in Arabic).

ARABIC

- (11) *'al-_ulaam-aani* *l-musiiqiyy-aani* *llað-aani*
 the-boy-DU.NOM the-musical-DU.NOM REL-DU.NOM 'the

two boy musicians whom Cyrano sent'

A number of other strategies are found cross-linguistically, nearly all of them much more frequently than the relative pronoun strategy. For a comprehensive account, including various subtypes of the types listed below, reference may be made to Lehmann (1984). One possibility is "non-reduction", whereby the head noun appears as a full-fledged noun phrase within the relative clause, with two major subtypes. In the correlative subtype, as illustrated by Hindi sentence (12), the head noun appears as a full-fledged noun phrase in the relative clause and is taken up again at least by a pronoun or other pronominal element in the main clause; a more literal translation would be 'which man I was talking with, he will go to India tomorrow'.

HINDI

(12)

<i>Maīm</i>	<i>jīs</i>	<i>ādmī</i>	<i>se</i>	<i>bāt</i>	<i>kar</i>
I.DIR	which.SG.OBL	man	to	talk	do
<i>rahā</i>	<i>thā</i>	<i>vah</i>	<i>kal</i>		
PRG.SG.M	be.IPF.SG.M	that.DIR.SG	tomorrow		
<i>bhārat</i>	<i>jāegā.</i>				
India	go.FUR.M.SG				

'The man [to whom I was talking - will go to India tomorrow.'

The so-called internally headed relative clause is illustrated by example (13) from Maricopa, a Yuman language; here the head is represented by a full noun phrase inside the relative clause, and has no explicit representation in the matrix clause:

MARICOPA

- (13) [*Kw̃nho mvar ily-m-uuchash*]-*sh* *m'ily-k.*
 basket flour IN-2-PUT:NMZR-SBJ infested-REALIS
 'The flour [that you keep - in the basket] is infested.'

Very frequent cross-linguistically is the so-called pronoun-retention strategy, illustrated by Persian sentence (14), which could be translated more literally as 'the men that you had given the books to them', a construction type

that does occur on the periphery of standard colloquial English and is also usual in several other European languages and language varieties that are more removed from the pan-European literary-language tradition. In the pronoun-retention strategy, the position relativized is explicitly indicated by means of a personal pronoun:

PERSIAN

- (14) *mardbâi [ke ketâbbâ-râ be ânbâ dâde bud-id]*
 men that books-ACC to them given were-2SG
 'the men [that you had given the books to -]'

Finally, there is the gap strategy, in which there is no overt reference whatsoever to the head noun within the relative clause. Compare Japanese simple sentence (15) with the relative clause (16), the latter containing no overt direct object, although the verb *kau* 'buy' is transitive:

JAPANESE

- (15) *Gakusei ga hon o katta.*
 student NOM book ACC bought
 'The student bought the book.'

- (16) *[gakusei ga katta] hon*
 student NOM bought book
 'the book [that the student bought -]'

The gap strategy is, of course, also available as an option in English, as in the alternative translation of (16) as *the book the student bought*, with the structure *the book [the student bought -]*, using the same notation as was used above and as will be explained below. Even *the book [that the student bought -]* is arguably an instance of the gap strategy, at least if that is analyzed as a conjunction rather than as a proposed direct object relative pronoun.

2. Extraction and the fact-S construction

In generative grammar of the 1960s a particular way arose of analyzing relative clauses of the Standard Average European type, one that has been influential in the analysis of languages well beyond this type. This analysis can be referred to as an extraction analysis, as illustrated by English examples (17)-(18), the former being a simple clause, the latter the corresponding relative clause:

ENGLISH

(17) *I saw the man*(18) *the man [whom I saw -]*

Since in English direct objects normally follow their verb, (18) can be analyzed by assuming that at some level of representation the direct object does follow the verb, but that in the sentence as it actually appears in English that direct object has been moved to clause-initial position. This account, or any of a number of essentially similar but notationally different other accounts, nicely explains a number of features of (18). In English in general, transitive verbs require a following direct object, but although the verb *saw* in (18) is as transitive as in (17) it does not have a following direct object -because that direct object has been preposed within the relative clause. The accusative case marking on *whom* is readily explainable in that at some level of representation this noun phrase is the direct object of *saw*, and thus would be expected to stand in the accusative case. Under this analysis, we can say that the relative pronoun is extracted from the position that it would have occupied in a simple sentence.

Now, this analysis for languages like English turns out to have a major added advantage, one not, to my knowledge, discussed in traditional grammar (where the intuition behind extraction is already present) but one that has turned to have a major impact within and beyond generative grammar. If there is extraction in the above sense, then it is conceivable that there might be constraints on such extraction, and indeed it turns out that there are, as was first discussed in detail by Ross (1967/1988). Consider examples (19)-(24):

(19) *I think [that the man has left].*(20) **the man [who I think [that - has left]]*(21) *I think [the man has left].*(22) *the man [who I think [- has left]]*(23) *[John plays the flute] and [Mary sings madrigals].*(24) **the madrigals [which [John plays the flute] and [Mary sings -]]*

In (20), the attempt has been made to extract the subject of a subordinate clause that is introduced by an overt conjunction (complementizer), in this case *that*, and the result is ungrammatical. Thus, English has a constraint against this extraction. The fact that this is a syntactic constraint is strongly suggested by the grammaticality of (22), which differs from (20) only in that the conjunction *that* has been omitted; (22) is perfectly well-formed in English. Thus, slight syntactic differences can give rise to different grammaticality judgments when one variant

falls foul of a constraint whereas the other does not. Sentence (24) illustrates one of the most famous of Ross's original constraints, the so-called Coordinate Structure Constraint, whereby it is not possible to extract an element from only one conjunct of a coordinate construction.

Indeed, the evidence for such constraints in English is so strong that when apparent exceptions are found, the first line of attack is usually not to assume that the constraint is wrong, but to look for an alternative analysis for the corresponding simple sentence so that the constraint is no longer violated. English example (26) seems to be a violation of the Coordinate Structure Constraint, but is nonetheless grammatical:

ENGLISH

(25) *I went to the store and bought the book.*

(26) *the book [which I went to the store and bought -]*

Rather than assume that (26) is an acceptable violation of the Coordinate Structure Constraint, linguists have tended (though not quite without exception) to assume an alternative analysis of the construction of (25), with *go* and *V* as some kind of complementation rather than coordination.²

Extraction also underlies the notation that I have introduced above. The long dash indicates the position from which the relative pronoun is extracted, the so-called "gap". Square brackets mark off the relative clause. In examples from languages other than English, I have generally indicated the square brackets, but not the long dash, to avoid unwanted presuppositions about the analysis of relative clauses in those languages. In English examples, and also in the English translations of examples from other languages, I have indicated the gap by means of the long dash.

The extraction analysis also provides a solution for an important difference between relative clauses in English (and other European languages) and an apparently similar construction that has come to be called the "fact-S construction". Compare examples (27) and (28):

ENGLISH

(27) *the book [which/that I criticized -]*

(28) *the fact [that I criticized the book]*

(29) *The students know [that I criticized the book].*

Example (27) is a relative clause construction. It may be introduced by *that*, but crucially it can also be introduced by the unequivocal relative pronoun *which*, in preposed position within the relative clause. Moreover, the relative clause in (27) has a well-defined gap, since the transitive verb *criticize* in English normally requires an overt direct object. Example (28) differs from (27) in the two crucial respects that define the relative pronoun strategy. First, in (28) *that* cannot be replaced by *which* (*the fact which I criticized the book), or indeed by any other case-marked pronoun. Second, the subordinate clause in (29) contains no gap: *I criticized the book* is a perfectly complete and well-formed transitive clause in English. In fact, the structure of the subordinate clause in (28) exactly parallels the structure of the subordinate clause in (29), the only difference being that in (28) the whole noun phrase is headed by a noun, in (29) by a verb. Example (29) is a canonical example of complementation in English, the subordinate clause being complement of the verb *know*. Example (28) thus illustrates complementation where the subordinate clause is complement of a noun. Although (27) and (28) may look superficially similar, it is easy to show that differences between them far outweigh the similarities, with (28) linked by far more similarities to (29).

To conclude this part of the exposition, extraction provides a good account of a range of phenomena connected with relative clauses in English and other Standard Average European languages. But does it provide a good account of relative clauses in all languages? In section 3, I will suggest that it does not. But before embarking upon this demonstration, I want to emphasize that the research program that was initiated by Ross (1967/1988) has led to the uncovering of a vast amount of information about human language from a cross-linguistic perspective. Indeed, one of my generative colleagues once told me that he thought this was the major empirical contribution of generative grammar over the past thirty years. Moreover, while any criticism of this research program might be taken specifically as an attack on generative grammar, this is by no means an obvious conclusion. Indeed, much of the work on relative clauses from a typological perspective has been guided by similar considerations, including not only the major comprehensive survey of relative clause formation from a typological viewpoint (Lehmann 1984), but also earlier work in which I have myself been involved (Keenan & Comrie 1977).

3. Equivalent constructions in Japanese

3.1. Japanese relative clauses

I want now to turn to a contrast between relative clause formation in English and relative clause formation in Japanese. Japanese relative clauses were already introduced in section 1, examples (15)-(16), which are repeated below as (30)-(31):

JAPANESE

- (30) *Gakusei ga hon o katta.*
 student NOM book ACC bought)
 'The student bought the book.'
- (31) [*gakusei ga katta*] *hon*
 student NOM bought book)
 'the book [that the student bought-]'

I will review the major differences between relative clause formation in English and Japanese, insofar as they affect the differences that are relevant in this paper. (Other differences, such as the fact that the relative clause precedes the head noun in Japanese but follows it in English, are not directly relevant.) In Japanese, there is no relative pronoun, i.e. no pronoun that can be identified as extracted and case-marked. Indeed, in the Japanese relative clause as illustrated in (31) there is no pronoun at all. Is there a gap? The answer might seem to be affirmative, since the relative clause of (31) lacks a direct object, although the verb *kau* 'buy' is transitive. However, it turns out that a simple sentence like (32) is grammatical in Japanese, with omission of the direct object, indeed (32) is the normal Japanese translation of English 'the student bought it':

JAPANESE

- (32) (*Gakusei ga katta*)
 student NOM bought)
 'The student bought [it].'

More generally, noun phrases (and postpositional phrases) can be omitted in Japanese if they are recoverable from context, i.e. under much the same conditions as English uses personal pronouns. Thus it is not obvious that (31) in Japanese really does contain a gap, in the way in which its English translation equivalent does.

Given these two sets of differences between Japanese and English -the lack of relative pronoun in Japanese, and the lack of an obvious gap in Japanese -there are two ways in which one might proceed. First, one might assume that despite the superficial differences, Japanese is really just like English, with the exception that it does not show overtly the effects of extraction that are so salient in English. This is the line that has been adopted explicitly in most generative work on Japanese, with talk of "covert movement", and indeed at least implicitly in a fair amount of the non-generative, including typological, work as well. The second possibility would be to take the Japanese construction at face value, in other words to assume that these overt differences between Japanese and English reflect a fundamental difference in construction. A detailed analysis following the second of these alternatives is proposed by Matsumoto (1988; *in press*), and I will adopt it here and work through some of its consequences for the typology of relative clauses.

Matsumoto's analysis is that the structure of (31) is just what it appears to be. A clause, identical in basic structure to that of a simple sentence (cf. (32)),³ is attached to a head noun, as a noun-modifying clause. This is all that the syntax says about the construction. Of course, for the construction to make sense the speaker of Japanese has to be able to infer a plausible relation between the head noun and the modifying clause. Various strategies, primarily semantic and pragmatic, are discussed in more detail in Matsumoto's work, and to a certain extent below, and I will restrict myself here to outlining a possible solution in the case of (31). The verb *kau* 'buy' suggests a certain frame (in the sense of Fillmore (1982)), since the act of buying requires someone to do the buying, someone to do the selling, something to be bought, as well as money or some equivalent to facilitate the transaction, and of course a time and a place at which the action takes place. As a first approximation, we can say that the speaker of Japanese will look for a relation by trying to interpret the head of the relative clause as one of the missing elements from the scene. In example (31), the most plausible is the thing to be bought. Returning to the syntax, we note that on this analysis not only is there no relative pronoun, there is no extraction and, therefore, no gap. But if this analysis is correct, then a number of other questions arise. Matsumoto (1988; *in press*) answers many of them. I will just take up a couple of these questions here, including some that are discussed by Matsumoto and others that are not.

3.2 Japanese fact-S constructions

What, for instance, of the fact-S construction in Japanese? Japanese has a

translation equivalent of the English fact-S construction, as in (33):

JAPANESE

- (33) [*gakusei ga hon o katta*] *zizitu*
 student NOM book ACC bought fact)
 'the fact [that the student bought the book]'

What is the syntactic structure of (33)? At least at first sight, it looks like a noun-modifying clause that has the same basic clause structure as a simple sentence (in fact, (30) above) preceding a head noun. Let us assume that this is the syntactic structure of (33). Then the construction is in fact identical to that for relative clauses, or more accurately: Japanese has a single construction, henceforth the noun-modifying clause construction, that subsumes the translation equivalents of both English relative clauses and English fact-S constructions. The interpretation problems faced by the Japanese speaker encountering (33) are in principle the same as in the case of (31), although the details are different, in particular in that the head noun *zizitu* 'fact' has as part of its frame the content of the fact in question, and this is precisely the interpretation that can plausibly be assigned.

So far, I have suggested that (33) can be analyzed as an instance of the same construction as in (31), in contrast to the situation in English, where the fact-S construction (28) differs markedly from the relative clause construction (27). But one might still wonder whether (33) cannot also be plausibly analyzed as an instance of complementation. In fact, it turns out that it cannot. Apart from a small number of highly idiomatized constructions that have inherited Old Japanese complementation constructions unchanged, Modern Japanese does not allow a clausal complement of a verb simply to end in the lexical verb (see, for instance, Horie (in press)). The two most common means of clausal complementation in Japanese with a verb as head are exemplified in (34) and (35):

JAPANESE

- (34) [*Gakusei ga hon o katta*] *no o sir-ana-katta*.
 Student NOM book ACC bought HEAD ACC know-NEG-PST
 'I didn't know that the student had bought the book.'
- (35) [*Gakusei ga hon o katta*] *koto o sir-ana-katta*.
 Student NOM book ACC bought thing ACC know-NEG-PST
 '[I] didn't know that the student had bought the book.'

The easier of these two examples to explain is (35). The verb 'know' takes as its immediate direct object the noun *koto* '(abstract) thing'. This noun, with rather abstract semantics, is preceded by the noun-modifying clause indicated by the square brackets. Crucially, clausal complementation with a verb in Japanese is parasitic on the noun-modifying clause construction already illustrated in (33), in other words in order to get to the construction of (35) one has to pass through the construction of (33). This is precisely the opposite of English, where the construction of (28) is parasitic on that of (29). Sentence (34), incidentally, has essentially the same structure as (35), except that here the head is even more abstract, being indicated by an element no lacking or virtually lacking lexical semantics.

If (33) and (31) are instances of the same construction, one might assume that there would be other instances of noun-modifying clauses that would receive interpretations distinct from either relative clauses or fact-S constructions. And this is indeed the case, as can be seen in (36), whose most natural English translation is with a gerundive nominalization joined to the head noun by the preposition *of*. In Japanese, there is simply a head noun and a clause modifying it.

- (36) [*dareka ga doa o tataku*] *oto*
 Someone NOM door ACC knock sound
 'the noise of someone knocking at the door'

The frame of the lexical item *oto* 'sound' -a sound must be the sound of something- facilitates filling in the relation between head noun and modifying clause.

3.3 Extraction and constraints in Japanese

If the above analysis of Japanese noun-modifying clauses is on the right lines, in particular for the analysis of those that are translation equivalents of English relative clauses, then a simple prediction is made. If indeed Japanese lacks extraction, then there can be no constraints on extraction in Japanese. The test of this conclusion seems straightforward: all one has to do is to see if the kinds of constraints that apply in languages like English -see examples (19)-(24)- also apply in Japanese or not. The data turn out to be less than immediately unequivocal, for the following reason. Given that, on the analysis adopted here, semantics and pragmatics play such an important role in the interpretation and acceptance of Japanese noun-modifying clause constructions, it is quite possible

that a particular construction might be judged unacceptable for reasons that have nothing to do with violations of alleged constraints on extraction, but with semantic or pragmatic problems. Some examples discussed by Haig (1996) will illustrate this point. Consider first the English translations of (37)-(38):

JAPANESE

- (37) [*Inu o katte ita*] *kodomo ga sindesimatta*.
 dog ACC keeping was child NOM died
 'The child [that - was keeping the dog] died.'
- (38) ?[[*katte ita*] *kodomo ga sindesimatta*] *inu*
 Keeping was child NOM died dog
 '*the dog_i [that the child_j [that -_j was keeping -_i] died]'

What is attempted in (38) is to relativize on a noun phrase, namely the dog, that is already part of a relative clause. The result is a totally ungrammatical construction. Indeed, the result is so ungrammatical that English speakers are even at a loss to know what the construction is meant to mean: typically, they cannot decide whether it is the child or the dog that died. In much of the generative literature on Japanese, it has been assumed that Japanese does have extraction and that this extraction can be constrained by restrictions similar to those proposed for English. And, as Haig notes, examples like (38) have sometimes been cited in this literature as ungrammatical, although there is certainly no unanimity on this judgment, and the question mark I have assigned to (38) should be interpreted to indicate a fairly wide range of disagreement among generative grammarians who have discussed this structure.

Suppose, however, that we improve the pragmatics of (38) by making it not an example involving some unknown child and some unknown dog, but relate it rather to one of the most famous stories in Japanese culture, namely that of the faithful dog Hachik_ who went each evening to Shibuya station to meet his master coming home from work. One day Hachik_'s master died at work, but Hachik_ was so loyal that he kept coming to the station each evening until his own death. Hachik_ is so famous that he even has his statue outside Shibuya station; few dogs can claim comparable status. In (39)-(40), in particular the crucial sentence (40), it is clear to anyone brought up in Japanese culture that this is about Hachik_, and this added information -in particular, the ability to identify the dog (Hachik_) and the person (Hachik_'s master)- means that the example is fully acceptable to speakers of Japanese, even those who have some hesitation about (38),

JAPANESE

- (39) [*Inu o kagawaigatte ita*] *bito ga nakunatta*.
 dog ACC keeping was person NOM died
 'The person [who - was keeping the dog] died.'
- (40) [*Kawaigatte ita*] *bito ga nakunatta*] *inu ga maiban*
 Keeping was person NOM died dog NOM every evening
eki made kainusi o mukae ni kita.
 station to master ACC greet to came
 '*The dog_i [that the person_j [who-_j was keeping -_i] died] came to the
 station every evening to greet his master.'

In other words, simply changing the pragmatics of a construction without changing its syntax can make it more acceptable, strongly suggesting that whatever is wrong with (38) is not syntactic.

3.4. Evidence from causative constructions

In fact, one can even go a stage further. It is not excluded that a particular instance of the noun-modifying clause construction might be ungrammatical for syntactic reasons, provided only that these syntactic reasons are not the direct result of violation of an alleged constraint on extraction. An obvious example in Japanese would be the constraint that postpositions cannot be stranded (in contrast to English, where prepositions can be stranded). Thus, in Japanese if a noun phrase is missing but its postposition is overt, the result will necessarily be ungrammatical, although this has nothing to do with extraction. Even a simple sentence with a missing noun phrase and an overt postposition is ungrammatical.

A more intricate example is provided by the following data concerning the causative construction in Japanese. In a Japanese causative construction using a lexically transitive verb, like *taku* 'cook' (used primarily of cooking rice), the causee (person made to carry out the action, in this case of cooking) must take the postposition *ni*, which among its other functions includes that of dative, as in (41). Although causees can sometimes be marked by the accusative particle *o*, in particular when the causative is of a lexically intransitive verb, this is not possible with a lexically transitive verb, as is shown by the ungrammaticality of (42):

JAPANESE

- (41) *Taroo ga Hanako ni mesi o tak-ase-ta.*
 Taro NOM Hanako to rice ACC cook-CAUS-PST
 'Taro made/let Hanako cook the rice.'
- (42) **Taroo ga Hanako o mesi o tak-ase-ta.*
 Taro NOM Hanako ACC rice ACC cook-CAUS-PST
 'Taro made/let Hanako cook the rice.'

There is general agreement in the literature that the ungrammaticality of (42) has something to do with a constraint against having two o-marked (or accusative-marked) noun phrases in the same clause-statements of the constraint vary in their details, and the area of controversy is not relevant to our present discussion; one early statement is given in Kuroda (1978). Note that this constraint has nothing to do with extraction. Now, suppose we want to relativize on *mesi* 'the rice' in (42). The grammatical result is as in (43), where the causee Hanako is marked by the postposition *ni*, while (44), using *o* to mark the causee, is ungrammatical:⁴

JAPANESE

- (43) [*Taroo ga Hanako ni tak-ase-ta*] *mesi*
 Taro NOM Hanako to cook-CAUS-PST rice
 'the rice [that Taro made Hanako cook -]'
- (44) **[Taroo ga Hanako o tak-ase-ta] mesi*
 Taro NOM Hanako ACC to cook-CAUS-PST rice
 'the rice [that Taro made Hanako cook -]'

Now, in the noun-modifying clause of (44), there is only one overt o-marked noun phrase, so it is not clear that the double-o constraint is violated. This might seem to be a nice piece of evidence in favor of the extraction approach. Because if the direct object of the subordinate clause of (44) is extracted, then there is necessarily a level of representation at which it is overt, and this level of representation will violate the double-o constraint.

It turns out, however, that this conclusion is too hasty. Bear in mind that one of the conditions on the noun-modifying clause construction in Japanese, quite independent of extraction or constraints on extraction, is that the basic clause structure of the noun-modifying clause must be a permissible

basic clause structure for a simple sentence in Japanese. The simple sentence corresponding to the subordinate clause in Japanese is as in (45), and (45) is just as ungrammatical as is (44):

JAPANESE

- (45) **Taroo ga Hanako o tak-ase-ta*
 Taro NOM Hanako ACC to cook-CAUS-PST
 'Taro made Hanako cook it.'

In other words, whether or not there is extraction in the analysis of (44), (44) will be excluded for the simple reason that (45) is excluded, and the ungrammaticality of (44) provides no evidence for the existence of extraction specific to the analysis of noun-modifying clauses in Japanese.

3.5. Evidence from benefactive constructions

The causative examples illustrated a case where, despite first appearances, one does not need to make an appeal to extraction, although they do not actually provide evidence against extraction. Indeed, one might say of all the data presented so far that they show that extraction is not necessary for the analysis of Japanese translation equivalents of English relative clauses, but they do not show that the extraction analysis is actually wrong. Of course, the fact that the analysis advocated here, avoiding extraction, provides a coherent account of an extensive set of data is a good argument in favor of that analysis. But still, it would be nice to have data that speak directly against the extraction analysis. Evidence of this kind seems to be forthcoming from the benefactive construction in Japanese.⁵

In Japanese, in order to express a beneficiary it is necessary to add to the sentence a verb of giving, such as *yaru* (past tense *yatta*), expressing the beneficiary as indirect object (with the particle *ni*) of this verb of giving, as in (46), which might be expressed somewhat more literally in English as 'buying this book, Taro gave to Hanako':

JAPANESE

- (46) *Taroo wa Hanako ni hon o katte yatta.*

Taro TOP Hanako to book ACC buying gave
 'Taro bought a book for Hanako.'

The idea of benefactive is thus in a sense expressed twice: by the verb 'give' and by the *ni*-marked noun phrase. In some cases, it is possible to imply a beneficiary by using a 'give' verb, but not to express it overtly with a *ni*-phrase. This happens, for instance, when one wishes to speak of dying for someone, with the intransitive verb *sinu* 'die'. Compare (47) and (48):

JAPANESE

- (47) **Hoken o takusan kakete, Taroo wa tuma ni sinde yatta.*
 insurance ACC much taking Taro TOP wife to dying gave
 'After taking out a lot of insurance, Taro died for his wife.'
- (48) *Kanozō no koto o kangaete, Taroo wa sinde yatta.*
 girlfriend GEN matter ACC thinking Taro TOP dying gave
 'Thinking about his girlfriend, Taro died for [her].'

In (47), the beneficiary is expressed overtly, and the result is ungrammatical. In (48) the indirect object is omitted, although its referent is readily retrievable from the first clause, and the result is grammatical. Now what about the corresponding relative clauses? If relativization in Japanese involves extraction, then the relative clause relativizing on the indirect object of the second clause of (48) should be ungrammatical, as there is no indirect object to be relativized. If the translation equivalent of 'the girl for whom Taro died' is simply a noun-modifying clause construction in which the noun-modifying clause has to have the same basic clause structure as a well-formed simple sentence, then the result should be grammatical in Japanese, just as the simple sentence comprising the main clause of (48) is grammatical. In fact, the result is grammatical, as shown in (49):

- (49) [*Taroo ga sinde yatta*] *kanozō*
 Taro NOM dying gave girlfriend
 'the girlfriend [for whom Taro died -]'

This provides a particularly strong piece of evidence in favor of the

analysis advocated here, using a single noun-modifying clause construction to cover Japanese translation equivalents of English relative clauses, fact-S constructions, etc. It argues against the extraction analysis.

4. Further typological considerations

So far, I have tried to argue that there are (at least) two types of languages with respect to the expression of the content of what in English is expressed by means of the relative clause construction. There are languages like English, which have a specific relative clause construction which involves extraction from the position relativized; this means that there can be constraints on such extraction, and that the relative clause construction will necessarily differ from other constructions (such as the fact-S construction) that do not involve extraction. And there are languages like Japanese, which have a single construction attaching a modifying clause to a head noun to form a noun phrase, lacking extraction (and thus, necessarily, constraints on extraction), and at least offering the possibility of extending this construction beyond translation equivalents of English relative clauses. So far, the discussion has largely been limited to these two languages. A number of further questions arise, including areal typological questions -how widespread among the world's languages is each of these two types?-and more general typological questions, for instance: how does this distinction correlate with other classifications of relative clause constructions, such as the classification including relative pronoun, pronoun-retention, and gap types.

In Comrie (1997) I try to give very preliminary answers to these questions, which I will summarize briefly here. Let me take the areal question first. At least a *prima facie* case can be made for the "Japanese" type in a number of languages in the following areas: East Asia (Japanese, Korean, Ainu, Chinese); Southeast Asia (Thai, Khmer -for which see also Comrie & Horie 1992-though apparently not Vietnamese); the Dravidian languages of South Asia (but not the Indo-Aryan languages, which typically have the correlative type illustrated in (12)); some Turkic languages. (I return to Turkic languages below.) I suspect that Haruai, a Papuan language of Highland New Guinea on which I did field work in the mid-1980s, also has a unified noun-modifying clause construction, although unfortunately at the

time of my fieldwork I was not aware of the importance of this typological distinction and did not test it in detail. And I emphasize that, for most of the languages discussed, it is primarily a case of *prima facie* evidence, since the relevant phenomena in these languages have not been investigated to the same extent as, for instance, in Japanese; and the history of the recent investigation of these phenomena in Japanese shows that great care must be taken, especially in disentangling the contribution of grammatical and of pragmatic factors to native-speaker judgments of acceptability. It would certainly not surprise if many other languages that have been described as using the gap strategy to form relative clauses turned out to behave similarly, although as I will show below one cannot simply equate what has hitherto been called the gap strategy with the unified noun-modifying clause construction.

As I suggested in section 1, the relative pronoun strategy, which provides the clearest evidence for a distinct relative clause construction with constraints on extraction, seems to be very limited in its areal distribution, effectively to European languages (and even more specifically to the standard varieties of the major European languages). It may thus reflect a rare type, not only in the theoretically irrelevant sense that few languages happen to belong to this type, but in the more relevant sense that this is a type that one would expect to be rare across a balanced sample of human languages reflecting the human language potential. The fact that linguists have tried to analyze translation equivalents of European relative clauses in the same way as they have analyzed relative clauses in European languages is thus a particularly striking way in which linguistic theory has been influenced by the accidental (and in this case probably rare) properties of European languages. Moreover, the influence of certain European languages (especially English, Spanish, and Russian) on languages of other parts of the world, especially languages that are in close symbiosis with a dominant European language, means that the hitherto rare European type of relative clause has been borrowed and is being borrowed into a number of other languages. I will cite just one example from Ewenki, a Tungusic language of Siberia which has been developed as a written language since the 1930s in Russia. The traditional Ewenki translation equivalent of European (including Russian) relative clauses uses the gap strategy, with a participial form of the verb in the subordinate clause, which

may precede or follow its head noun, as in (50)-(51):

EWENKI

- (50) Tār gūl_{ə-lə} ɪ-rə-n [girkɪ-w inj_{ə-rɪ-lə-n}].
 he house-into go-AOR-3SG friend-my live-PTCPL-into-3SG
 'He went into the house [where my friend lives -].'

- (51) Bəjəl il-la gūləkən-dū [ɪduk asɪ yū-rə-n].
 Men stand-AOR.3PL hut-at whence woman come.out-AOR -
 3SG
 'The men stood by the hut [from which the woman had come -].'

However, under the influence of Russian, the European type of relative clause has started appearing in written Ewenki, as in (52), which follows the Russian type even to the extent of using an interrogative pronoun as the relative pronoun:

EWENKI

- (52) Amakān yulədū-wun gūlasəg ičəw-rə-n [anti-wa bu
 soon in.front-1PL village appear-AOR-3SG which-ACC we
 ə-čə-wun sə-rə].
 NEG-PAST-1PL know-AOR
 'Soon in front of us appeared a village [which we did not know -].'

(The Ewenki examples are taken from Kolesnikova (1966).) In an important contribution to the study of language contact, Johanson (1992) suggests that two factors are important in promoting the borrowing of a construction into another language: the construction may be structurally "attractive", and thus likely to be borrowed even in the absence of strong cultural pressure, or it may be that the prime motivation for its borrowing is cultural pressure from the dominant language; I suspect that the spread of the European-type relative clause is an instance of the latter.

But let me now turn to the more general typological question, namely the correlation between the "European"/"Japanese"-type distinction and other classifications of relative-clause forming strategies. The relative-pronoun strategy must, by logical necessity, be of the "European" type, since it clearly

involves extraction, may therefore have constraints on extraction, and thus has relative clauses distinct from other constructions. Other strategies, in particular the gap strategy, may be of the "Japanese" type, but there is no logical necessity that they should be so, and indeed it turns out that some languages with the gap strategy behave otherwise more or less like European languages, while others behave like Japanese. A good example of the former is provided by the alternative construction in English with a zero relativizer, as in (53) (and compare (1)):

ENGLISH

(53) *the boy [I saw -]*

Although the construction illustrated in (53) uses the gap strategy, with no reference to the head noun within the relative clause, it is subject to the same kinds of constraints as are relative-pronoun relative clauses in English; compare (54)-(56) with (20), (22), (24) above:

ENGLISH

(54) **the man [I think [that - has left]]*

(55) *the man [I think [- has left]]*

(56) **the madrigals [[John plays the flute] and [Mary sings -]]*

An even more interesting comparison is to be found within the Turkic language family, as discussed in somewhat more detail in Comrie (1997). The Turkic languages are genetically very closely related, with fairly high degrees of mutual intelligibility. I will here compare Karachay-Balkar, a Turkic language of the northern Caucasus, and Turkish, the official language of Turkey. Karachay-Balkar has the features of the "Japanese" type, as illustrated in (57)-(61):

KARACHAY-BALKAR

(57) [*kitab-i al-γan*] *oquwču*
 book-ACC buy-PRT student
 'the student who bought the book'

(58) [*oquwču al-γan kitap*]

student buy-PRT book
'the book that the student bought'

- (59) [tüb-ün-dän qara-sa-η, bōrkü-η tüš-gän] esgi
base-3SG-from look up-CND-2SG cap-2SG fall-PRT old
narat tārāk
fir tree
'the old fir tree that when you look up from the bottom your hat falls off'⁶

- (60) [prezident kel-gän] hāpar
president come-PRT news
'the news that the president has come'

- (61) [et biš-gän] iyis
meat cook-PRT smell
'the smell of meat cooking'

Examples (57)-(58) show that the same construction is used for translation equivalents of English sentences relativizing on different constituents (here, subject versus direct object). Example (59) shows that it is possible to relativize on constituents that would typically be inaccessible in languages using the European type of relative clause. Example (60) shows that the same construction as in (57)-(59) can be used as a translation equivalent of the fact-S construction. Example (61) shows that this is a general noun-modifying clause construction, as (61) parallels Japanese example (36). Comparative Turkish material is provided in (62)-(63):⁷

TURKISH

- (62) [kitab-ı al-an] öğrenci
book-ACC buy-PRT student
'the student who bought the book'

- (63) [öğrenci-nin al-dığ-ı] kitap
student-GEN buy-NMZ-3SG book
'the book which the student bought'

- (64) [cumhurbaşkanı-nın gel-diğ-i] haber-i
 president-GEN come-NMZ-3SG news-3SG
 'the news that the president has come'

As (61) and (62) indicate, Turkish uses different constructions -a participle versus a nominalization- depending (very approximately) on whether relativization is on a subject or a non-subject, i.e. the position relativized must be syntactically identified in the formation of the Turkish relative clause. Example (64) shows that the fact-S construction in Turkish, though similar to that for relativizing on a non-subject (cf. (63)), is not identical, as the fact-S construction requires a possessive suffix on the head noun. A literal translation of Karachay-Balkar (61) is simply ungrammatical in Turkish.

The comparison between Karachay-Balkar and Turkish, two genetically closely related languages with superficially similar syntax for clauses that modify head nouns, shows that great care must be taken in assigning particular languages using the gap strategy to the "European" or "Japanese" type. I suspect that the same will turn out to be true for languages using the pronoun-retention strategy -where again the translation equivalent of a relative clause has the basic structure of an independent clause -although this will have to be the subject of future research.

5. Conclusions

The points that I have tried to make in this article can be interpreted at two different levels. First, the article can be interpreted as a contribution to the typology of relative clauses. Much of the typology of relative clauses, like much of the formal grammatical study of relative clauses, has been concerned with such concepts as extraction or accessibility. If the general line of this article is correct, then these notions are simply inapplicable to languages of the "Japanese" type. This does not, of course, mean, that they are irrelevant for other languages, indeed I believe that they remain crucially relevant for languages of other types. But an initial typological dichotomy must be drawn between languages subject to extraction constraints and languages not subject

to extraction constraints. Only within the latter group does it then make sense to talk about different kinds of extraction constraints, etc.

This article can also be interpreted at a more general methodological level, as a partial answer to the question: How similar to and how different from each other can languages be? In the study of relative clauses and their translation equivalents, there has been a tendency, both in formal and in typological studies, to assume that relative clauses will necessarily have a structure that is amenable to the same kinds of analytical tools as have worked for European languages. I have tried to show that "relative clauses" in Japanese not only look different from their European translation equivalents, but also are different. But I emphasize that in order to establish these differences, it was important to look in detail at the properties of relative clauses and related constructions in the various languages. And section 4 illustrated that it is still possible to have superficially similar languages, namely Karachay-Balkar and Turkish, which turn out to differ on precisely those parameters that are central to the present investigation. On a somewhat more specific point, I have tried to show that the relative contribution of grammatical and pragmatic considerations to acceptability judgments is different as between the two kinds of languages. While some linguists tell me that they are interested in grammar but not pragmatics (or vice versa), this article shows that linguistic phenomena do not come conveniently labeled as grammatical or pragmatic; whatever one's main focus of interest in investigating language, one must always be open to the possibility that an apparently grammatical phenomenon will turn out to be pragmatic, or vice versa.

To put the point more generally: Formal grammarians must be just as sensitive as typologists to the differences among languages. And typologists cannot afford to be any less scrupulous analytically than their formal grammarian colleagues. So the overall moral of this article is that what we need is careful analysis that is sensitive to the differences across languages.

Notes

1. An earlier version of this article was presented to the Plenary Meeting of the European Science Foundation Project on Language Typology (EUROTYP) in Le Bichenberg, France, in

1994, under the title "How exotic are European languages? -the case of the relative clause"; this title obviously takes off from Dahl (1990). Other versions were presented to a number of audiences in Europe, Asia, and North America, in addition to the Seminario Internacional de Tipología Lingüística in Granada. I am grateful to all those who have contributed to discussions of this paper. The following abbreviations are used: ABS-absolutive, ACC-accusative, CAUS-causative, DIR-direct, DU-dual, ERG-ergative, FUT-future, GEN-genitive, IPF-imperfect, M-masculine, NEG-negative, NMZR-nominalizer, NOM-nominative, PL-plural, PRG-progressive, PRS-present, PST-past, REL-relative, SBJ-subject, SG-singular, TOP-topic.

2. Compare the unexceptional extraction in (i):

(i) the book [that I went to the store to buy -].

Note, however, that (26) cannot simply be analyzed as a stylistic variant of (i), since they differ in meaning. Sentence (26) entails that I bought the book, while (i) does not.

3. In Modern Japanese, the relative clause in (31) can indeed be used without change as a simple sentence, namely (32), though this is rare cross-linguistically, and even in Modern Japanese there are differences irrelevant to our present concerns, namely that a simple sentence like (32) can end with a sentence-final particle, whereas a relative clause as in (31), like most subordinate clauses, cannot. In most languages that otherwise follow the "Japanese pattern", the verb in the relative clause is non-finite, as is indeed the case in such languages for many kinds of subordinate clause. This was also, incidentally, the case in Old Japanese; Modern Japanese has lost this particular finite/non-finite distinction in verb morphology. These differences are not relevant to our present concerns.

4. Given that the Japanese verb *taku* is virtually restricted to cooking rice, it is not even easy to get the somewhat gruesome, cannibalistic interpretation under which Hanako is the object of cooking, i.e. something like 'the rice with which Taro had Hanako cooked'.

5. Examples (46)-(48) are taken from Shibatani (1994).

6. Note that the English translation of (59) given in the text is of very dubious acceptability, especially in the written language.

7. I am grateful to Jaklin Kornfilt for encouraging me to convince myself that Turkish is not of the "Japanese" type. Lars Johanson and the Turkish students at the University of Mainz helped me well beyond the call of duty with the Turkic material.

References

- Comrie, Bernard and Horie, Kaoru. 1995. 'Complement clauses versus relative clauses: some Khmer evidence'. In W. Abraham, T. Givón, S.

- Thompson (eds.), *Discourse grammar and typology*, 65-75. Amsterdam: John Benjamins.
- Comrie, Bernard. 1997. 'Attributive clauses in Asian languages: towards an areal typology'. In Winfried Boeder, Christoph Schroeder (eds): *Sprache in Raum und Zeit. In memoriam Johannes Bechert*, 000-000. Tübingen: Gunter Narr.
- Dahl, Östen. 1990. 'Standard Average European as an exotic language'. In J. Bechert, G. Bernini, C. Buridant (eds.), *Toward a typology of European languages*, 3-8. Berlin: Mouton de Gruyter.
- Fillmore, Charles J. 1982. 'Frame semantics'. In: *Linguistic Society of Korea* (ed.), *Linguistics in the Morning Calm*, 111-138. Seoul: Hanshin.
- Haig, John H. 1996. 'Subacency and Japanese grammar: a functional account'. *Studies in Language* 20:53-92.
- Horie, Kaoru. In press. 'The history of complementation in Japanese: a cognitive perspective'. To appear in *Linguistics*.
- Johanson, Lars. 1992. *Strukturelle Faktoren in türkischen Sprachkontakten* [Structural factors in Turkic language contacts]. Stuttgart: Franz Steiner.
- Keenan, Edward L. and Bernard Comrie. 1977. 'Noun phrase accessibility and universal grammar'. *Linguistic Inquiry* 8:63-99.
- Kolesnikova, V.D. 1966. *Sintaksis Švenkijskogo jazyka* [Syntax of the Ewenki language]. Moscow-Leningrad: "Nauka".
- Kuroda, S.-Y. 1978. 'Case marking, canonical sentence patterns, and counter equi in Japanese (a preliminary survey)'. In J. Hinds and I. Howard (eds.), *Problems in Japanese syntax and semantics*. Tokyo: Kaitakusha.
- Lehmann, Christian. 1984. *Der Relativsatz: Typologie seiner Strukturen, Theorie seiner Funktionen, Kompendium seiner Grammatik* [The relative clause: typology of its structures, theory of its functions, compendium of its grammar]. Tübingen: Gunter Narr.
- Matsumoto, Yoshiko. 1988. 'Semantics and pragmatics of noun-modifying constructions in Japanese'. *Berkeley Linguistic Society* 14: 166-175.
- . In press. *Noun-modifying constructions in Japanese: a frame semantics approach*. Amsterdam: John Benjamins.
- Ross, John R. 1967. *Constraints on variables in syntax*. PhD dissertation, MIT, Cambridge MA. [Published as: *Infinite syntax*. Norwood MA, 1986: Ablex.]

- Shibatani, Masayoshi. 1994. 'Benefactive constructions: Toward a formal analysis of the schema-based cognitive approach'. Paper presented at the Language Typology Workshop, Institute for the Study of Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.
- Shimomiya, Tadao. 1978. *Zur Typologie des Georgischen (verglichen mit dem Indogermanischem), mit einem Exkurs zur Sprachbundtheorie* [On the typology of Georgian (compared with Indo-European), with an excursus on sprachbund theory]. Tokyo: Gakushuin University Press.